



TITLE:

地域一貫性評価指標の性能評価 (種々のモデルのための漸近展開とそれらに関連する話題)

AUTHOR(S):

吉田, 顕臣; 高橋, 史朗; 竹内, 正弘

CITATION:

吉田, 顕臣 ...[et al]. 地域一貫性評価指標の性能評価 (種々のモデルのための漸近展開とそれらに関連する話題). 数理解析研究所講究録 2013, 1860: 42-57

ISSUE DATE:

2013-11

URL:

<http://hdl.handle.net/2433/195295>

RIGHT:

地域一貫性評価指標の性能評価

北里大学大学院医学研究科 臨床薬学研究・教育センター 臨床統計学
Department of Clinical Medicine (Biostatistics), Graduate School of
Pharmaceutical Sciences, Kitasato University

吉田 顕臣(Akiomi Yoshida)

高橋 史朗(Fumiaki Takashi)

竹内 正弘(Masahiro Takeuchi)

1.序論

近年、医薬品開発の国際化に伴い、東アジア諸国でも国際共同試験が数多く実施されるようになってきた。国際共同試験では地域を併合した薬効に主たる興味があるが、外的・内的要因による薬効の違いについても詳細な吟味を必要とする。

現在は一貫性評価の指標として広く用いられている方法は PMDA が例示した方法^[1]で、「全地域の薬効差の点推定値がすべて同じ符号をもつ」場合に一貫性があるとする評価方法(PM2)である。前述のとおり地域を併合した薬効に興味があるため、各地域の結果は試験集団の部分集団である。特に東アジア民族などを一つの集団として捉えられないような状況では、全地域から十分な推定精度をもつ症例数を組み入れられることは少ない。症例数の少ない地域がある状況では一貫性の証明は難しくなる。これを示す先行研究として Hung ら^[2]が全地域を併合した際の 2 群比較で p 値が 5% で有意となる状況において、4 地域のうち 1 地域あるいは 2 地域で薬効差が逆転する確率および PM2 で一貫性を示せない確率が計算されている(表 1)。この結果より PM2 は真の薬効が示せているにも関わらず、およそ 50~60%は一貫性なしと判断してしまう問題がある。症例数の少ない地域が含まれることで一貫性なしとなる確率は大きくなる傾向がある。また、この研究ではデータの質は全地域で等しいと仮定して、計算を行っているがデータの質が各地域で異なるような状況では一貫性なしとなる確率はさらに高くなると考えられる。このように PM2 は一貫性評価を行う際に症例数およびデータの質の影響を受けるという問題がある。そのため、症例数とデータの質の影響を受けない新たな一貫性評価指標が必要である。

表 1

症例数設定	1 地域で薬効差が逆転する確率	2 地域で薬効差が逆転する確率	一貫性なしとなる確率
1:1:1:1	0.38	0.11	0.49
2:1:3:4	0.40	0.13	0.53
1:1:1:7	0.43	0.17	0.60

我々は新たな一貫性評価指標の提案のために国際共同試験のデータが階層構造をもつことに注目した。Shih^[3] は地域における効果のバラツキを表現し、地域の効果を推定するために経験ベイズ(EB) 法や階層ベイズモデルの利用を推奨している。これらの方法は興味のある地域の薬効の評価に他の地域のデータを利用することができるため、症例数の少ない地域がある状況で利用可能な有用な方法であると考えられた。しかしながら一貫性評価指標の提案はなされていない。

そこで本研究では、Efron^[4]が EB 法の枠組みで提案している類似性の指標 R (興味のあるデータが他のデータと同じ分布から抽出されたかを評価) を国際共同試験に適応させることを提案する。提案方法の性能を評価するために、モンテカルロ・シミュレーションを通して既存方法^[5] と比較検討を行うことを目的とする。

次項からは一貫性評価に用いる評価指標について解説し、研究で実施したシミュレーションについて記述する。3 章で結果を示し、4 章で考察を行う。

2.方法

2.1.1 類似性の指標 R

本研究では、Efron^[4]によって提案された経験ベイズ法に基づく類似性の指標を評価することにより、国際共同試験における薬効の一貫性を評価することを考える。経験ベイズ法とは階層ベイズ法の一つであり、以下のようなデータ構造を持つ場合に適用する解析方法である。

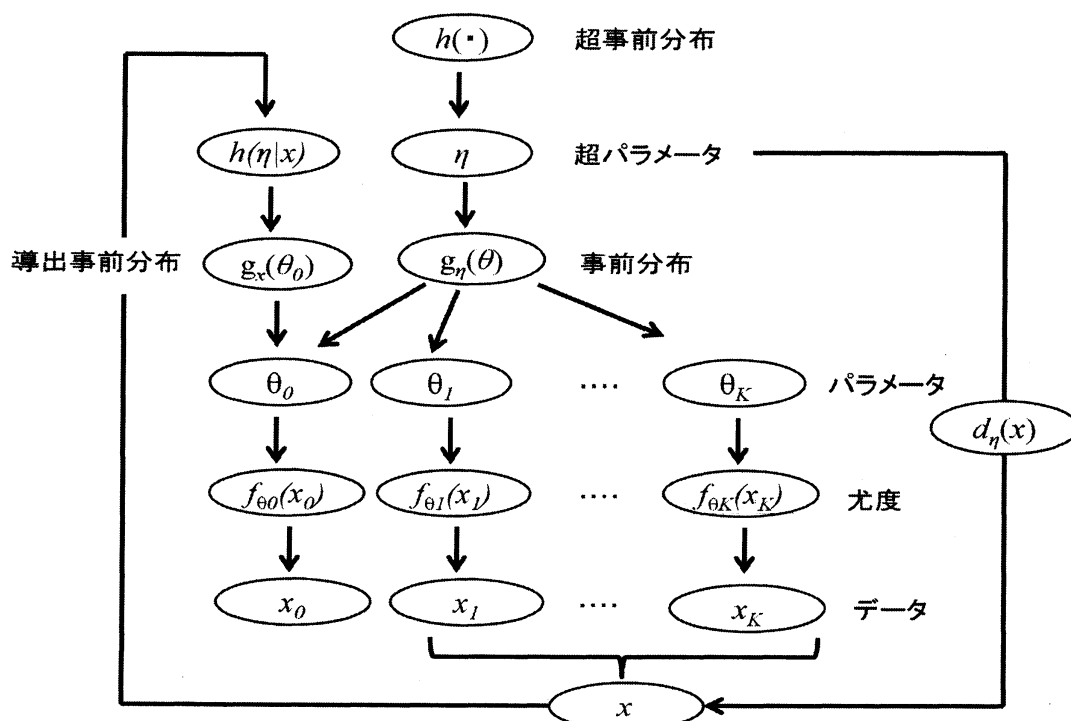


図 1 国際共同試験のデータ構造

この構造を国際共同試験のデータの構造に適応すると、 $\theta = (\theta_0, \dots, \theta_K)$ は $(K+1)$ 地域の薬効差の母平均を表している。本研究では θ_0 を興味のある地域、例えば日本の薬効差の母平均として推定を行い、これが他の地域の母平均と一貫性があるか評価することを目的とする。

まず、階層ベイズ法による推定方法について説明する。通常のベイズ法ではパラメータがある事前分布 $g_\eta(\theta)$ に従うと仮定するが、その際事前分布のパラメータ（超パラメータ η ）は解析者が主観的に決める。

$$\theta_0, \theta_1, \dots, \theta_K \stackrel{\text{i.i.d}}{\sim} g_\eta(\theta)$$

階層ベイズ法では、超パラメータ η にも分布（超事前分布 $h(\cdot)$ ）を仮定し、2段階以上の変動を表す。これによって、地域や国間の違い（外的要因、内的要因）と地域内、国内での変動（観測誤差）を分離して考えることができる。

興味のある地域の薬効差データ x_0 を除いた他の地域のデータ $(x = (x_1, \dots, x_K))$ の周辺標本密度関数 $d_\eta(x)$ は、未知パラメータ θ を積分することにより、以下の式で与えられる。

$$d_\eta(x) = \int g_\eta(\theta) f_\theta(x) d\theta$$

ここで、 $g_\eta(\theta) = \prod_{k=1}^K g_\eta(\theta_k)$ 、 $f_\theta(x) = \prod_{k=1}^K f_\theta(x_k)$ である。これは興味のある地域以外の薬効差に対する事後確率を掛け合わせたものとなる。これをもとに他の地域の情報で更新した超事前分布 $h(\eta|x)$ は、以下の式で与えられる。

$$h(\eta|x) = c h(\eta) d_\eta(x)$$

ここで c は正の基準化定数である。この超事前分布を用いて興味のある地域の真の薬効差 θ_0 に対する導出事前分布 $g_x(\theta_0)$ が以下の式で与えられる。

$$g_x(\theta_0) = \int h(\eta|x) g_\eta(\theta_0) d\eta$$

この導出事前分布を用いて推測を行うのが階層ベイズ法であり、事後分布は

$$p_\eta(\theta_0|x_0) = c g_\eta(\theta_0) L_0(\theta_0)$$

で表される。ここで、 $L_0(\theta_0) \equiv c f_{\theta_0}(x_0)$ とする。この事後分布をもとに推定量 $\widehat{\theta_0}$ を求める。以上が階層ベイズ法の手順となる。

一方、経験ベイズ法の場合は、周辺密度 $d_\eta(x)$ を求めるところまでは同じであるが、事前分布のパラメータ η を最尤法 (MLE) に基づいて推定する。つまり、

$$\hat{\eta} = \operatorname{argmax}_{\eta} \{d_{\eta}(x)\}$$

となるように推定する。この式は $d_{\eta}(x)$ が最大になるような η を求めることを意味する。先ほどの事後分布の式に $\hat{\eta}$ を代入したものが経験ベイズ法による事後分布となる。

$$p_{\hat{\eta}}(\theta_0|x_0) = c g_{\hat{\eta}}(\theta_0) L_0(\theta_0)$$

次に本研究で用いる類似性の指標 R について説明する。まず、図 1 国際共同試験のデータ構造が仮定できるとして、 $g_{\eta}(\theta_0)$ に 2 通りの事前分布 g_A, g_B を考える (θ_0 がそれぞれ確率 $h_A, h_B (= 1 - h_A)$ で従うと仮定する)。 g_A は経験ベイズ法で推定される事前分布を表し、 g_B は g_A ではない別の事前分布、例えば無情報事前分布を想定する。無情報事前分布とは事前情報として利用できる情報がないことを意味する事前分布であり、分散が無限大の正規分布や Jeffrey's の事前分布 (Fisher 情報行列の行列式に比例) などがある。2 通りの事前分布それぞれで事後周辺密度は、以下の式で与えられる。

$$d_A(x_0) = \int g_A(\theta_0) f_{\theta_0}(x_0) d\theta_0$$

$$d_B(x_0) = \int g_B(\theta_0) f_{\theta_0}(x_0) d\theta_0$$

θ_0 の事前分布を g_A と g_B の混合分布と考えると、周辺標本密度関数は次のように書くことができる。

$$d(x_0) \equiv h_A d_A(x_0) + h_B d_B(x_0)$$

g_A を選択する事後確率と g_B を選択する事後確率は、それぞれ次のように与えられる。

$$h_A(x_0) = h_A d_A(x_0) / d(x_0)$$

$$h_B(x_0) = h_B d_B(x_0) / d(x_0)$$

このとき、類似性の指標 R はこれらの事後確率の比

$$R = \frac{h_A(x_0)}{h_B(x_0)} = \frac{h_A}{h_B} R(x_0)$$

$$R(x_0) = \frac{\int g_A(\theta_0) f_{\theta_0}(x_0) d\theta_0}{\int g_B(\theta_0) f_{\theta_0}(x_0) d\theta_0}$$

と定義される。事後確率の比であるので R は Bayes Factor であり、この値が基準値を超えた場

合、経験ベイズ法による事前分布がより適していることを意味する。そのため他の地域の情報を利用することが有用であることを意味し、興味のある地域の薬効差は他の地域の薬効差と似ていと解釈する。

2.1.2 提案法

本研究の提案法は先述の類似性の指標 R を一貫性評価指標にするために全地域を統合したデータでの 2 標本検定の結果と組み合わせるものである。一般的な国際共同試験では全地域を併合した際に、実薬群と対照群で 2 群比較を実施した場合、有意な差があると考えられる。一貫性の評価は第 3 相試験で行われるものであり、対照群よりも一定以上の効果が見込まれることが期待できる。そのため、地域間の違いを考慮しなければ有意な差があることが一般的であると考えられる。この点に注目し、提案法は以下の 2 つの条件を満たした場合に一貫性があると判断する。

- (1) 全データにおける 2 標本 t 検定で有意な差が示される。
- (2) R が 3 もしくは 10 以上となる。

R の基準値の設定については R がベイズファクターであることから、ベイズファクターの基準値を定めた論文(Kass & Raftery^[6]) の値を参照した(表 2)。

表 2

ベイズファクター	一貫性ありの支持の程度
$BF_{01} < 1$	対立仮説を支持
$1 < BF_{01} < 3$	支持しない
$3 < BF_{01} < 20$	支持する
$20 < BF_{01} < 150$	強く支持する
$150 < BF_{01}$	非常に強く支持する

2.2.シミュレーションによる検討

本研究では、結果変数を連続変数とし、混合効果モデルに従う実薬群と対照群の乱数を生成し、提案法の一貫性証明確率を他の手法と比較検討することを目的とする。比較検討する手法は PMDA が例示する方法、Cochran-Q 統計量、Gail-Simon の質的検定とした。

混合効果モデル：固定効果（治療群など）と変量効果（地域など）が混在するモデル

$$y_{ktj} = \beta_0 + \beta_{1t}X_{ktj} + b_{1kt}Z_{ktj} + \epsilon_{ktj}$$

y_{ktj} : 地域 k の治療 t の被験者 j の応答変数 ($k = 1, \dots, K$, $t = A \text{ or } C$, $j = 1, \dots, n_k$)

β_0 : 固定効果の切片項

β_{1t} : 固定効果の傾き項

b_{1kt} : 傾き項の変量効果 ($b_{1kt} \sim N(0, \sigma_{kt}^2)$)

X_{ktj} : 固定効果の説明変数(治療の有無, $X_{kAj} = 1, X_{kCj} = -1$)

Z_{ktj} : 変量効果の説明変数

ϵ_{ktj} : 誤差項 ($\epsilon_{ktj} \sim N(0, \sigma^2)$)

$$y_{ktj} \sim N(\beta_0 + \beta_{1t}X_{ktj}, \sigma_{kt}^2 + \sigma^2)$$

シミュレーションは、1 群あたり症例数 264 例(総症例数 528 例)の二群比較を 4 地域で行う無作為化比較試験を想定した。本研究では、実薬群と対照群の症例数は等しいと仮定した。症例数設定の根拠は真の薬効差が 0.2、実薬群、対照群の分散がいずれも 0.5 となるエフェクトサイズを想定し、有意水準 5%、検出力 90%を達成する例数として設定した。

シミュレーション設定条件：全 30 通り

症例数：全地域で等しい場合(1 : 1 : 1 : 1)、1 地域だけ少ない場合(1 : 3 : 3 : 3)の 2 通り

データの質 (誤差項の分散 σ^2) : 全地域で等しい場合(1 : 1 : 1 : 1)、1 地域だけ質が良い場合(0.5 : 1 : 1 : 1)、1 地域だけ質が悪い場合(2 : 1 : 1 : 1)の 3 通り

真の薬効差：5 通りでいずれも一貫性あり

表 3

症例数	データの質 (分散 σ^2)	真の薬効差 ($\beta_{1A}X_{kA} - \beta_{1C}X_{kC}$)
1 : 1 : 1 : 1	1 : 1 : 1 : 1	0.2 : 0.2 : 0.2 : 0.2 (一貫性あり)
1 : 3 : 3 : 3	0.5 : 1 : 1 : 1	0.2 : 0.1 : 0.2 : 0.2 (一貫性あり)
	2 : 1 : 1 : 1	0.1 : 0.2 : 0.2 : 0.2 (一貫性あり)
		0.1 : 0.05 : 0.2 : 0.2 (一貫性あり)
		0.05 : 0.1 : 0.2 : 0.2 (一貫性あり)

混合効果モデルにおける誤差分散 σ^2 と変量効果の分散 σ_{kt}^2 の比率は 9999 : 1 とした。

2.2.1 結果変数発生のメカニズム

結果変数の発生には 2 段階の乱数発生を行う。シミュレーション i 回目の地域 k の実薬群、対照群の母平均をそれぞれ θ_{kAi} , θ_{kCi} とする。

平均が上述の 8 パターンの真の薬効差 $\mu_{kA} - \mu_{kC}$ 、分散が変量効果の分散項 $0.5 * \tau$ の正規分布からの乱数として θ_{kAi} と θ_{kCi} を 10000 個ずつ発生させる($\tau = 0.0001$)。

$$\theta_{kAi} \sim N(\beta_1 X_{kA}, 0.5 * \tau)$$

$$\theta_{kCi} \sim N(\beta_1 X_{kC}, 0.5 * \tau)$$

i 回目のシミュレーションの実薬群と対照群の結果変数 $x_{kAi} | \theta_{kAi}, x_{kCi} | \theta_{kCi}$ は次のようにあらわすことができる。

$$x_{kAij} | \theta_{kAi} \sim N(\theta_{kAi}, 0.5 * (1 - \tau))$$

$$x_{kCij} | \theta_{kCi} \sim N(\theta_{kCi}, 0.5 * (1 - \tau))$$

ここでやっていることは、最初に発生させた θ_{kAi} と θ_{kCi} を平均、データの質 $0.5 * (1 - \tau)$ を分散とする正規分布に従う乱数を地域 k に対応する例数分発生させることである。これが各地域の試験データとなる。

2.2.2 シミュレーションで比較する一貫性評価指標

- ・PMDA が例示する方法 (PM2)
- ・Cochran-Q 統計量(Cochran)
- ・Gail-Simon の質的検定(GS)

これらの手法は点推定値そのものもしくは計算された検定統計量によって評価を行う。詳しくは付録を参照。

- ・提案法

提案法については類似性の指標 R を計算する際に用いる無情報事前分布と事前選択確率 ha, hb の設定 について説明する。

無情報事前分布には分散を無限大に近付けた (10^{10} の分散) 正規分布 ($N(0, 10^{10})$) を用いた。分散を無限に近付けることで、情報がないことを表現している

全ての地域が同じ事前分布 $g_A(\theta)$ に従っている確率 ha は 0.95、無情報事前分布に従っている確率は 0.05 と設定した。またベイズファクターの基準値は 3 を用いた。

3.結果

3.1.例数一定で分散を変化させたシミュレーション結果

3.1.1.例数一定、等分散の場合（一貫性ありの場合）

例数一定、等分散の場合のシミュレーション結果を示す。

それぞれの手法の一貫性証明確率を計算し、結果を棒グラフにまとめた（図 2 各手法の一貫性証明確率（例数一定、等分散））。グラフの縦軸が一貫性証明確率、グラフの横軸は真の薬効差の設定を表している。

一貫性証明確率は $GS > PM2 > \text{提案法} > \text{Cochran}$ の順で高かった。薬効差が小さい地域の増加によって一貫性証明確率が低下する傾向が全手法でみられた。しかしながら、GS の一貫性証明確率は薬効差によってほとんど変化はなかった。

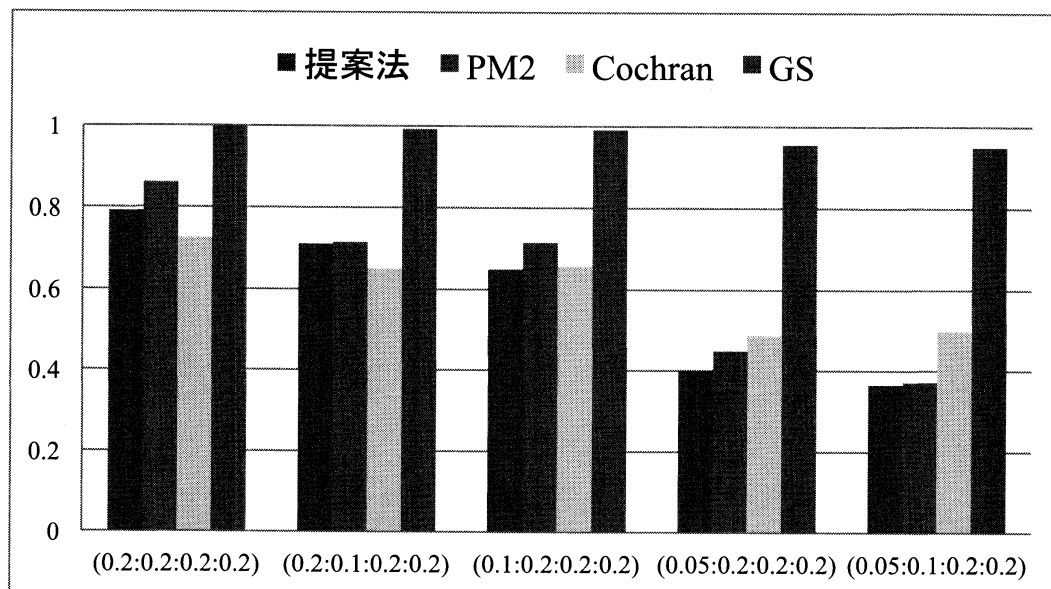


図 2 各手法の一貫性証明確率（例数一定、等分散）

3.1.2. 例数一定、分散が半分の場合（一貫性）

3.1.1 の状況から興味のある地域のデータの質を良くした場合のシミュレーション結果を示す。3.1.1 と比べると提案法は薬効差の小さい地域を含む状況では一貫性証明確率は低下した。PM2 の一貫性証明確率は薬効差の小さい地域を含まない状況では一貫性証明確率が高くなった。Cochran の一貫性証明確率は全体的に低下した。GS には大きな変化は見られなかった。

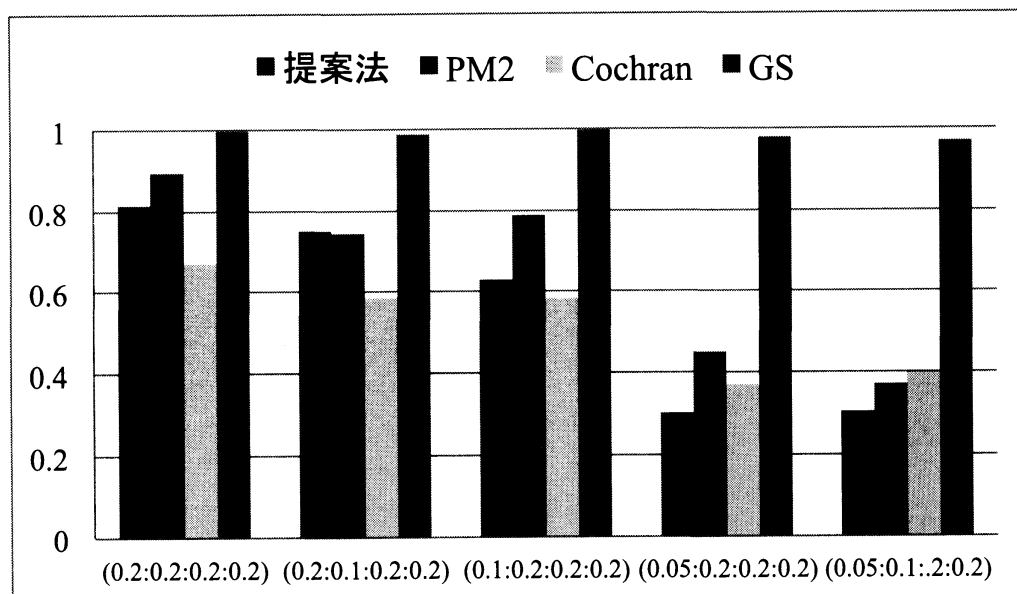


図 3 各手法の一貫性証明確率（例数一定、分散半分）

3.1.3. 例数一定、分散 2 倍の場合

3.1.1 の状況から興味のある地域のデータの質を悪くした場合のシミュレーション結果を示す。提案法、PM2 は 3.1.1 に比べ薬効差の小さい地域を含まない状況では一貫性証明確率は低下したが、薬効差の小さい地域を含む状況では提案法の一貫性証明確率が高くなった。Cochran の一貫性証明確率は全体的に高くなった。GS には大きな変化は見られなかった。

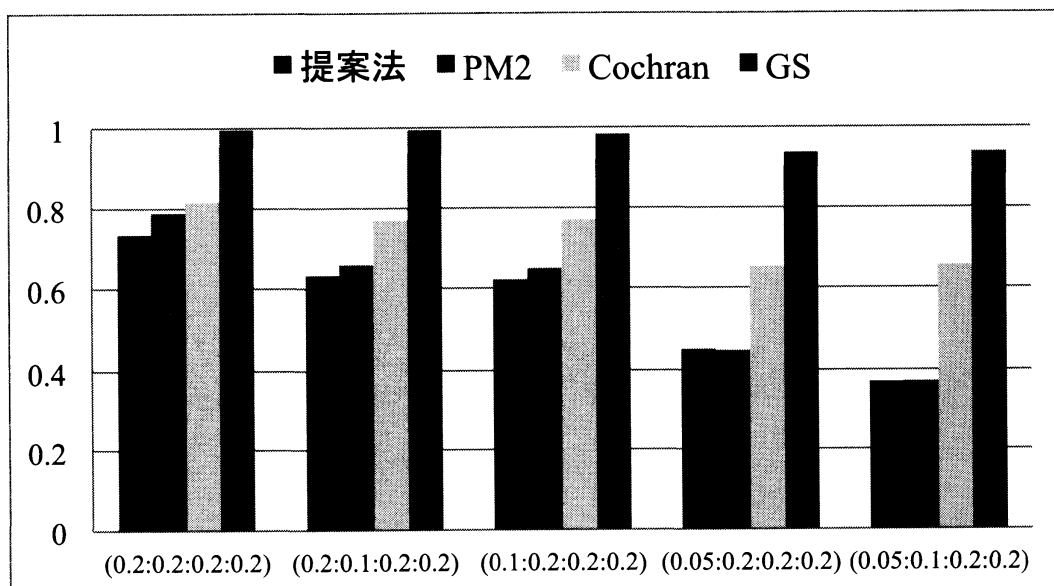


図 4 各手法の一貫性証明確率（例数一定、分散 2 倍）

3.2. 例数 1:3:3:3 で分散を変化させたシミュレーション結果

3.2.1 例数 1:3:3:3、等分散の場合

3.1.3 までは興味のある地域のデータの質を変化させたシミュレーション結果を示した。ここからは興味のある地域の例数を少なくした上で分散を変化させたシミュレーション結果を示す。例数が一定の3.1.1 と比べると提案法と Cochran は薬効差が小さい地域を含む状況では一貫性証明確率は高くなった。PM2 は逆に薬効差が小さい地域を含まない状況で一貫性証明確率が低下した。GS には大きな変化は見られなかった。

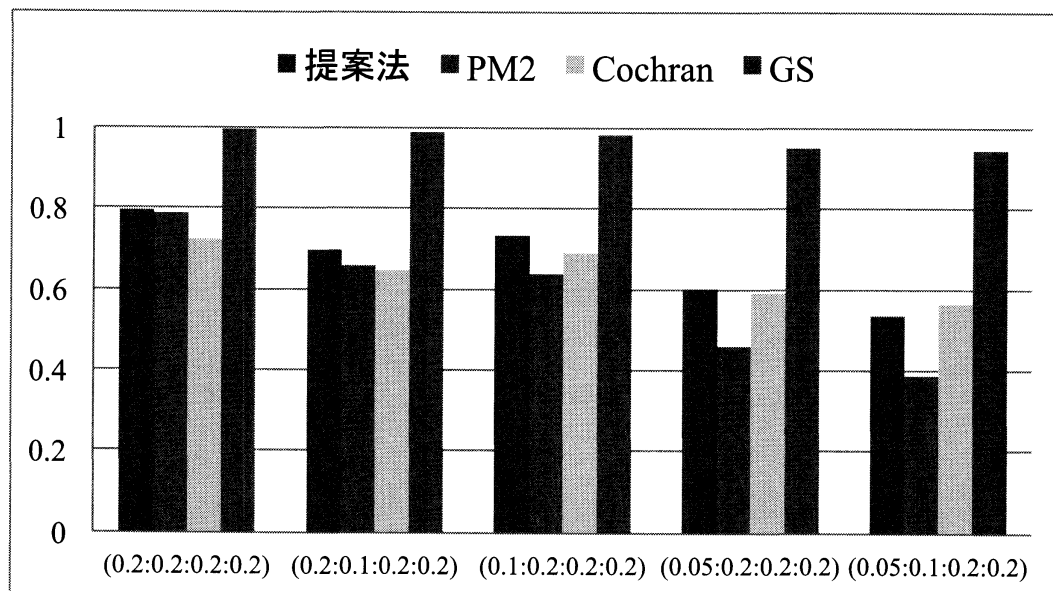


図 5 各手法の一貫性証明確率（例数 1 : 3 : 3 : 3、等分散）

3.2.2 例数 1 : 3 : 3 : 3、分散半分の場合

3.2.1 から興味のある地域のデータの質を良くした結果を示す。3.2.1 との比較つまり分散による影響は例数が同じであった 3.1.1 と 3.1.2 の比較結果とほぼ同じであった。

例数の違いによる影響を評価するため、3.1.2 と比較を行う。3.1.2 と比べると提案法の一貫性証明確率は薬効差が小さい地域を含む状況で高くなった。PM2 はほとんど変化はないが、興味のある地域の薬効差のみがやや小さい状況（グラフの 3 番目）で一貫性証明確率が高くなった。Cochran は全体的に高くなった。GS もほとんど変化は見られなかったが、グラフの 2 番目の状況での一貫性証明確率が小さくなった。

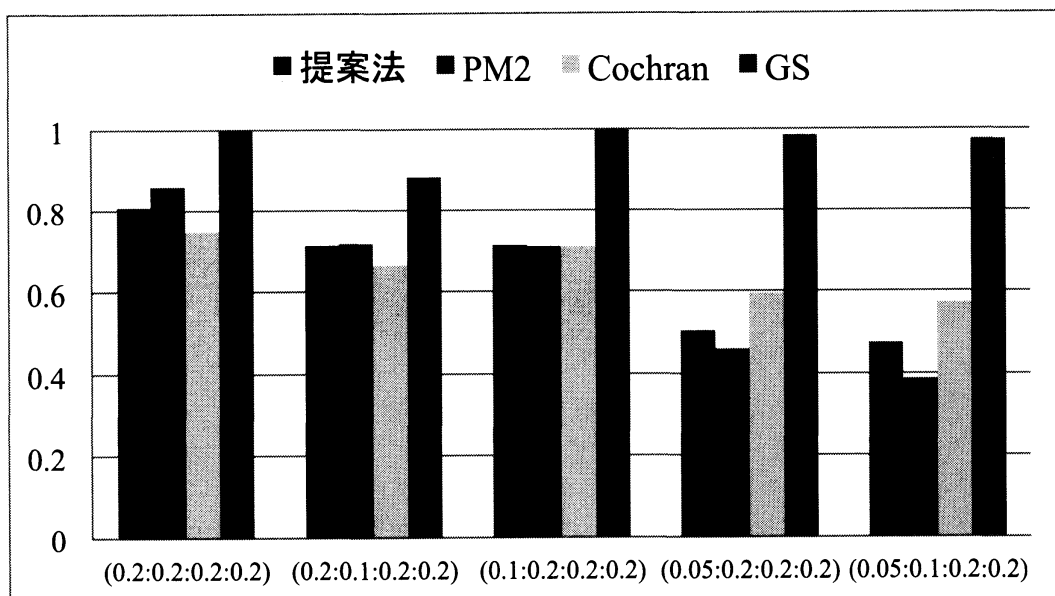


図 6 各手法の一貫性証明確率（例数 1 : 3 : 3 : 3、分散半分）

3.2.3 例数 1 : 3 : 3 : 3、分散 2 倍の場合

3.2.1 から興味のある地域のデータの質を悪くした結果を示す。3.2.1 と比べると提案法は薬効差の小さい地域を含む状況では一貫性証明確率が高くなり、薬効差の小さい地域を含まない状況では大きな変化はなかった。PM2 は薬効差の小さい地域を含まない状況では一貫性証明確率が低下した。Cochran と GS は大きな変化はなかった。例数が同じであった 3.1.1 と 3.1.3 の比較とは異なる比較結果となった。

例数の影響を評価するため 3.1.3 と比較する。3.1.3 と比べると提案法の一貫性証明確率は全体的に高くなり、Cochran は逆に全体的に低くなった。PM2 は薬効差の小さい地域を含まない状況では一貫性証明確率が低下した。GS はほとんど変化しなかった。

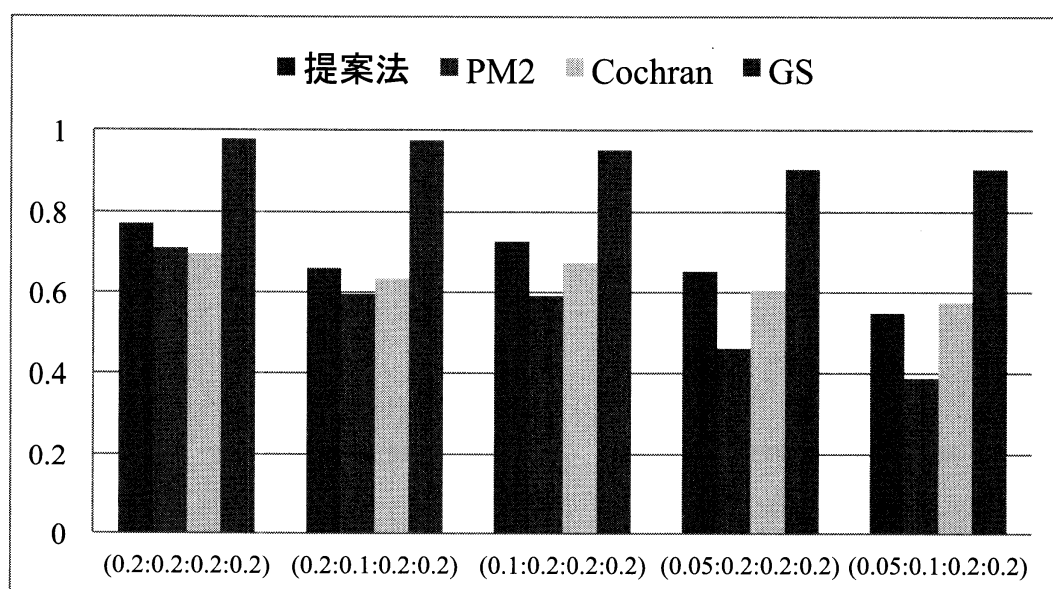


図 7 各手法の一貫性証明確率（例数 1 : 3 : 3 : 3、分散 2 倍）

4. 考察

本研究では、一貫性評価指標として Efron の提案した類似性の指標 R と全地域を併合した際の 2 標本検定を組み合わせた新たな一貫性評価指標を提案し、既存の一貫性評価指標とシミュレーションによって比較検討した。シミュレーションの結果、提案法は例数が全地域で等しく、分散が全て等しいような理想的な状況では PMDA の例示する方法 2 よりも一貫証明確率が低くなった。しかしながら、例数が少ない地域が含まれ、分散が全地域で等しくないようなより現実的な状況では提案法の一貫性証明確率はあまり大きく変動せず、PM2 の一貫性証明確率が低下したため、結果として PM2 よりも提案法の一貫性証明確率が高いという結果になった。このことから、理想的な試験データを得ることができるのであれば PM2 の方が一貫性を示しやすいが、現実的には提案法の方が一貫性を示しやすい。

Cochran の Q 統計量は Efron の提案した方法とほぼ同様の挙動を示したが、薬効差の小さい地域を含まない状況では提案法の方が一貫性証明確率は高く、薬効差の小さい地域を含む状況では

Cochran の Q 統計量の一貫性証明確率が高くなるという傾向がみられた。このことから薬効の小さい地域を含むか含まないかで評価指標を使い分けることが望ましいと考える。

Gail-Simon の質的交互作用検定については薬効の大きさ、例数、分散によって一貫性証明確率に大きな変化は見られず、どのような状況でも一貫性証明確率が高い手法であることが分かった。しかしながら、薬効が小さい状況でも一貫性証明確率が高いため、実際に薬効が反転している地域が含まれていた場合にも一貫性を示してしまう可能性がある。他の手法でも同じことが言えるが、Gail-Simon の質的交互作用検定では確率が非常に高く、一貫性評価を誤りやすいので注意が必要である。

結論としては現実的なデータで一貫性評価を行う場合には Gail-Simon の質的交互作用検定、提案法、Cochran の Q 統計量、PM2 の順で一貫性を示しやすいが、Gail-Simon の質的交互作用検定は一貫性がない状況でも高い確率で一貫性を示してしまう恐れがある。一貫性評価を誤る可能性を考慮し、提案法による一貫性評価を推奨する。

本研究では、興味のある地域の分散と例数のみを変動させているため、まだ現実的なシミュレーションを実施できていない。今後は興味のある地域以外の例数、分散を変化させてシミュレーションを実施し、各手法の性能をより詳細に評価する必要があると考える。本研究で用いたシミュレーションプログラムは現在もデバッグを実施中であり、正確なシミュレーションを実施して評価を行う必要がある。

今後は Bayman らが提案した質的交互作用検定をベイズ流に拡張した方法^[9]および小谷らによる Bayman 法の拡張法^[9]との比較を実施する予定である。また、2013 年 Huang ら (Pharm. Stat 2013;12:59-64)^[10]は、興味のある国の薬効差の事前分布に、薬効がない状態と他国と同じ状態の分布の混合分布を想定し、薬効差の事後分布がゼロ以上となる確率を計算する経験ベイズ法を提案した。今後はこの方法とも性能比較を行う必要があると考える。

5. APPENDIX

5.1 PMDA が例示する方法

独立行政法人医薬品医療機器総合機構（PMDA）は「国際共同治験に関する基本的考え方について」の 6.で国際共同試験に組み入れる日本人の症例数設計について以下の 2 つの方法を記述している。

方法 1: 対照群と治験薬群での群間差を δ 、その場合の全集団での群間差を δ_{all} 、日本人集団における群間差を δ_{Japan} とすると、 $\delta_{Japan}/\delta_{all} > \pi$ が成立するような確率が 80 %以上となるように日本人症例数を設定する。 π については、適切な値を設定する必要があるが、一般的には 0.5 以上の値をとることが推奨される。この方法では、日本人症例数を最小にしようとすると、全体での症例数が増加し、全体での症例数を最小にしようとすると日本人症例数が増加するという関係が認められる。

方法 2: 全集団における対照群と治験薬群での群間差を δ_{all} 、例えば 3 地域が試験に参加し、

各地域での対照群と治験薬群での群間差をそれぞれ $\delta_1, \delta_2, \delta_3$ とすると、 $\delta_1, \delta_2, \delta_3$ が全て同様の傾向にあることを示す。例えば δ_{all} が正の値をとるとすると、 $\delta_1, \delta_2, \delta_3$ のいずれの値も 0 を上回る確率が 80 % 以上となるように症例数を設定する。この方法では、各地域から均等に症例数を集積した場合に、確率が高くなるという傾向があり、全体の症例数を変更することなく日本人症例数を検討することが可能であるが、日本人の構成比率が小さく、症例数が少ない場合に、地域間比較が十分に行えない場合があることに留意すべきである。

以上の方法 1 と方法 2 で示されている条件が試験結果で満たされた場合に一貫性があると判断するのが PMDA の例示する一貫性評価指標である。この指標は一貫性の評価に点推定値だけを用いるので、日本の症例数が少ない場合には、他の地域よりも推定精度が低下し、点推定値が真の値とは反対側に出てしまう危険性がある。

5.2 Cochran の Q 統計量

Cochran の Q 統計量はメタアナリシスで用いられる。地域 k の真の薬効差を δ_k とし、以下の仮説の検定を行う検定統計量である。

帰無仮説： $\delta_0 = \delta_1 = \dots = \delta_K$

対立仮説：帰無仮説の等号のいずれかが成り立たない

検定統計量は以下のように計算する。

(N_{Ak}, N_{Pk} は地域 k の実薬群、対照群の症例数, $\hat{\sigma}_{Ak}^2, \hat{\sigma}_{Pk}^2$ は実薬群、対照群の不偏分散を表す)

$$Q = \sum_{k=0}^K \frac{(\hat{\delta}_k - \hat{\delta})^2}{\left(\frac{1}{N_{Ak}} + \frac{1}{N_{Pk}}\right) \hat{\sigma}_k^2}$$

$$\hat{\sigma}_k^2 = \frac{(N_{Ak} - 1)\hat{\sigma}_{Ak}^2 + (N_{Pk} - 1)\hat{\sigma}_{Pk}^2}{N_{Ak} + N_{Pk} - 2}$$

統計量 Q は自由度 K のカイ 2 乗分布に従うため、有意水準 α のパーセント点を基準に検定を行う。この統計量は全地域の薬効差が同じかどうかを評価する統計量なので、一貫性評価では、判断を誤る可能性がある。例えば、全地域の薬効が 0 を超えていても、1 地域だけ薬効差が小さいなどの場合には一貫性を示しにくくなると考えられる。

5.3 Gail-Simon の質的検定

Gail-Simon の質的検定とは、全地域の薬効差がそれぞれ同じ性質を持っているかを評価する検定である。この検定の帰無仮説と対立仮説は以下のとおりである。

$H_0: \delta_k > 0$ すべての $k=1, \dots, K$ もしくは $\delta_k < 0$ すべての $k=1, \dots, K$

$H_1: H_0$ が成り立たない

つまり、すべての薬効差が正もしくは負であることを検定する方法である。検定統計量は以下の式で計算する。

$$Q^- = \sum_{k=1}^K \frac{\hat{\delta}_k^2}{\left(\frac{1}{N_{Ak}} + \frac{1}{N_{Pk}}\right) \hat{\sigma}_k^2} I(\hat{\delta}_k < 0), Q^+ = \sum_{k=1}^K \frac{\hat{\delta}_k^2}{\left(\frac{1}{N_{Ak}} + \frac{1}{N_{Pk}}\right) \hat{\sigma}_k^2} I(\hat{\delta}_k > 0)$$

薬効差が正の場合と負の場合で2通りの統計量を計算し、これらのうち小さいほうが基準値を超えた場合に帰無仮説を棄却し、一貫性がないと評価する。本研究では地域数 K を4とし、基準値は5.43とした。基準値はGailらの研究^[7]で計算されている数値を用いた。

この検定はすべての薬効差が正なのか負なのかを評価するため、一貫性の評価指標である。

6.参考文献

- [1] 独立行政法人 医薬品医療機器総合機構 「国際共同治験に関する基本的考え方について」
<http://www.pmda.go.jp/operations/notice/2007/file/0928010.pdf> (2013年1月30日アクセス)
- [2] H.M.James Hung et al. (2010). Consideration of regional difference in design and analysis of multi-regional trials, *Pharmaceutical Statistics*, **9**:173–178
- [3] W.J.Shih (2006). Bridging Studies → Global Development
<http://www.amstat.org/meetings/fdaworkshop/presentations/2006/Shih.ppt>, (2012年7月19日アクセス)..
- [4] B.Efron (1996). Empirical Bayes Methods for Combining Likelihoods, *Journal of the American Statistical Association*, **91**: 538–565.
- [5] J.Chen et al. (2010). Assessing consistent treatment effect in a multi regional clinical trial : a systematic review. *Pharmaceutical Statistics*, **9**:242–253
- [6] Kass RE, Raftery AE (1995) Bayes Factors, *Journal of the American Statistical Association*, **90**: 773–795.
- [7] S. Piantadosi and M.H.Gail. (1993). A Comparison of the power of two tests for qualitative interactions. *Statistics in Medicine*. **12**:1239–1248.

- [8] Bayman EO et al (2010) Detecting qualitative interaction: A Bayesian approach, *Statistics in Medicine*, 29: 173–178.
- [9] 小谷 基 (2011). 国際共同試験において治療効果の地域間一貫性を評価するベイズ流アプローチ.2011 年度 日本計量生物学会年会講演予稿集 p1–6.
- [10] Yufen Huang et.al (2013),An empirical Bayes approach to evaluation of results for a specific region in multiregional clinical trials, *Pharmaceutical Statistics*, 12:59–64